Structural information in two-dimensional patterns: Entropy convergence and excess entropy

David P. Feldman[1,2,*] and James P. Crutchfield[2,†]
[1]*College of the Atlantic, Bar Harbor, Maine 04609*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*
(Received 5 December 2002; published 19 May 2003)

We develop information-theoretic measures of spatial structure and pattern in more than one dimension. As is well known, the entropy density of a two-dimensional configuration can be efficiently and accurately estimated via a converging sequence of conditional entropies. We show that the manner in which these conditional entropies converge to their asymptotic value serves as a measure of global correlation and structure for spatial systems in any dimension. We compare and contrast entropy convergence with mutual-information and structure-factor techniques for quantifying and detecting spatial structure.

DOI: 10.1103/PhysRevE.67.051104 PACS number(s): 05.20.−y, 05.45.−a, 65.40.Gr, 89.75.Kd

I. INTRODUCTION

The past decade has seen considerable advances in our understanding of general ways to detect and quantify pattern in one-dimensional systems. This work is of intrinsic and broad interest, since it suggests different ways of viewing patterns and calls attention to some of the subtleties associated with pattern discovery and quantification [1], issues that—implicitly or explicitly—underlie much of the scientific enterprise.

Recently, these abstract measures of structural complexity or pattern played a key role in several applications in physics and dynamical systems. For example, there is a growing body of work that seeks to relate the structural complexity of a one-dimensional sequence to the difficulty one encounters when trying to learn or synchronize to the generating process [2–5]. Also, complexity measures have recently been used to characterize experimentally observed structures in a class of layered materials known as polytypes [6].

The successes in one dimension have not been paralleled by similar advances in two dimensions. Nonetheless, the development of a general measure of complexity—or pattern or structure—for two-dimensional systems is a long-standing goal. How is information shared, stored, and transmitted across a two-dimensional lattice to produce a given set of configurations? How can we quantitatively distinguish between different types of ordering or pattern in two dimensions? Though largely answered in one dimension, these questions are open in higher dimensions.

Some recent work in this area, motivated in part by the need to characterize complex interfaces in surface science and geology [7–13], has suggested a set of approaches to these questions that are similar in spirit to fractal dimensions, in the sense that these approaches involve coarse-graining variables and then monitoring the changes that result as the coarse-graining scale is modulated. One can also use a multifractal approach, also known as the singularity spectrum "$f(\alpha)$," the thermodynamic formalism, and the fluctuation spectrum; for reviews, see, e.g., Refs. [14–16]. All of these approaches can be applied to spatial structures, but they suffer several drawbacks. For one, they are not fully spatial, in the sense that their calculation requires one to discard spatial information. Second, they do not directly speak to the correlation present in a system; rather they are more measures of entropy, disorder, and inhomogeneity.

Other recent general approaches to pattern in two dimensions include the extension of the formal theory of computation [17] and an information-theoretic approach [18] somewhat similar in spirit to that which we develop below. See also Ref. [19].

In this work, we take a different approach to the question of pattern and structure in two spatial dimensions. Our starting point is the *excess entropy*, an information-theoretic measure of complexity that is commonly used and well understood in one dimension [2,20–29]. Our main goals are several fold. First, we introduce three ways to extend the definition of excess entropy to more than one dimension, noting that these extensions are not equivalent. Second, we report results of estimating two of these forms of excess entropy for a standard statistical mechanical system: the two-dimensional Ising model with nearest- and next-nearest-neighbor interactions on a square lattice. We show that these two forms of excess entropy are similar but not identical, that each is sensitive to the structural changes the system undergoes, and that they are able to distinguish between different patterns that have the same structure factors. Third, we discuss some of the subtleties and challenges associated with moving from a one- to a two-dimensional information-theoretic analysis of pattern and structure.

II. ENTROPY AND ENTROPY CONVERGENCE IN ONE DIMENSION

We begin by reviewing information-theoretic quantities applied to one-dimensional (1D) systems. This allows us to define quantities and to fix notation that will be useful in our discussion of two-dimensional (2D) information theory in the subsequent section.

Let X be a random variable that assumes the values $x \in \mathcal{X}$, where $\mathcal{X}$ is a finite set. We denote the probability that X assumes the particular value x by $\Pr(x)$. Likewise, let Y be a random variable that assumes the values $y \in \mathcal{Y}$. The *Shannon*

*Electronic address: dpf@santafe.edu
†Electronic address: chaos@santafe.edu

entropy of the random variable X is defined by

$$H[X] \equiv -\sum_{x \in \mathcal{X}} \Pr(x) \log_2 \Pr(x). \tag{1}$$

The entropy $H[X]$ measures the average uncertainty, in units of bits, associated with outcomes of X. The *conditional entropy* is defined by

$$H[X|Y] \equiv -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr(x,y) \log_2 \Pr(x|y) \tag{2}$$

and measures the average uncertainty associated with variable X, if we know the outcome of Y. Note that Eq. (2) is the average of $-\log_2 \Pr(x|y)$, averaged over the joint distribution $\Pr(x,y)$. Finally, the *mutual information* between X and Y is defined as

$$I[X;Y] \equiv H[X] - H[X|Y]. \tag{3}$$

Thus, Y carries information about X to the extent that knowledge of Y reduces one's average uncertainty about X. The above three definitions are all standard; for details, see, e.g., Ref. [30].

A. Block entropy and entropy density

We now examine the behavior of the Shannon entropy $H(L)$ of a sequence of L random variables $S^L = S_0 S_1 \cdots S_{L-1}$. The total Shannon entropy of length-L sequences—the *block entropy*—is defined by

$$H(L) = -\sum_{s^L \in \mathcal{A}^L} \Pr(s^L) \log_2 \Pr(s^L), \tag{4}$$

where $\Pr(s^L)$ is the probability of a particular L block s^L. Pictographically, we represent this as

$$H(L) \equiv H[\overleftarrow{\square\square\square\square}^{L}\overrightarrow{\square}]. \tag{5}$$

The *entropy density* is then defined as

$$h_\mu \equiv \lim_{L \to \infty} \frac{H(L)}{L}. \tag{6}$$

The above limit exists for all spatial-translation invariant systems [30]. Equations (6) and (4), together, are equivalent to the Gibbs entropy density. The entropy density h_μ can be reexpressed as the limit of a form of conditional entropy. To do so, we first define

$$h_\mu(L) \equiv H[S_L | S_{L-1} S_{L-2} \cdots S_1]. \tag{7}$$

In words, $h_\mu(L)$ is the entropy of a single spin conditioned on a block of $L-1$ adjacent spins. This can also be written graphically

$$h_\mu(L) = H[\boxtimes | \overleftarrow{\square\square\square\square}^{L-1}\overrightarrow{\square}]. \tag{8}$$

The pictogram on the right indicates that the entropy is conditioned on the $L-1$ spins directly to the right of the single *target spin* $\boxtimes$, with the bold vertical lines denoting the boundary where the target spin and spin block abut. One can then show that the entropy density defined in Eq. (6) can be written as

$$h_\mu = \lim_{L \to \infty} h_\mu(L). \tag{9}$$

For a proof that the limits in Eqs. (9) and (6) are equivalent, see Ref. [30]. As the block length L grows, the terms in Eq. (9) typically converge to h_μ much faster than those in Eq. (6). See, e.g., Ref. [31], and citations therein.

B. Excess entropy

The entropy density measures the randomness or unpredictability of the system; h_μ is the randomness that persists even after correlations over infinitely long blocks of variables are taken into account. A complementary quantity to the entropy density is the *excess entropy* E [20–29], which accounts for how the finite-L entropy density estimates $h_\mu(L)$ converge to their asymptotic value h_μ. For each L, the system appears more random than it actually is by an amount $h_\mu(L) - h_\mu$. Summing up these entropy-density overestimates gives us the *excess entropy*

$$E_{\mathrm{C}} \equiv \sum_{L=1}^{\infty} [h_\mu(L) - h_\mu]. \tag{10}$$

The excess entropy thus measures the amount of *apparent* randomness at small L values that is "explained away" by considering correlations over larger and larger blocks. The subscript in E_{C} indicates that this form of excess entropy is defined by considering how the entropy density *converges* to h_μ. Note that E_{C} can be infinite for systems with long-range correlations [2,22,28,29].

Another expression for the excess entropy is obtained by looking at the growth of the block entropy $H(L)$. By Eq. (6), we know that $H(L)$ typically grows linearly for large L. The excess entropy can be shown to be equal to the portion of $H(L)$ that is sublinear—E is the *subextensive* part. That is, the excess entropy is defined implicitly by

$$H(L) = E_{\mathrm{S}} + h_\mu L \quad \text{as} \quad L \to \infty. \tag{11}$$

Here, the subscript in E_{S} serves as a reminder that this expression for the excess entropy is the *subextensive* part of $H(L)$.

Finally, one can show [24,28] that the excess entropy is also equal to the mutual information between two adjacent semi-infinite blocks of variables:

$$E_{\mathrm{I}} = \lim_{L \to \infty} I[S_{-L} \cdots S_{-2} S_{-1}; S_0 S_1 \cdots S_{L-1}] \tag{12}$$

$$= \lim_{L \to \infty} I[\overleftarrow{\square\square\square\square}^{L}; \overrightarrow{\square\square\square\square}^{L}]. \tag{13}$$

The "I" in the subscript indicates that this expression for the excess entropy is given in terms of a mutual *information*.

Note that in the pictographic version, Eq. (13), the two semi-infinite blocks are understood to be adjacent, as indicated by the thick vertical lines.

The three different forms for the excess entropy—E_C, E_S, and E_I—given above are all equivalent in one dimension [24,28]. We represent these different forms with distinct symbols because they are *not* identical in two dimensions.

In the subsequent section, we compare our results for the excess entropies with various structure factors—standard quantities from statistical physics used to detect periodic structure. The definition of the structure factor begins with the two-spin correlation function

$$\Gamma_{ij} \equiv \langle (s_i - \langle s_i \rangle)(s_j - \langle s_j \rangle) \rangle \tag{14}$$

$$= \langle s_i s_j \rangle - \langle s \rangle^2, \tag{15}$$

where s_i and s_j denote the value of spins at different lattice coordinates. The second equality follows from the translation invariance of configurations. The angular brackets indicate a thermal expectation value. In 2D, we will be interested in spins that are separated horizontally or vertically, but not both. (In a scattering scenario, this corresponds to restricting ourselves to a situation in which the particles to be scattered are incident along a line parallel to one of the axes of the lattice.) We define $\Gamma(r)$ as the correlation function between two spins separated, horizontally or vertically, by r lattice sites:

$$\Gamma(r) \equiv \langle s_0 s_r \rangle - \langle s \rangle^2. \tag{16}$$

The *structure factor*, then, is the discrete Fourier transform of the correlation functions:

$$S(p) = \sum_{r=1}^{\infty} \cos\left(\frac{2\pi r}{p}\right) \Gamma(r). \tag{17}$$

If the correlation function has a strong period-p component, then $S(p)$ is large; if not, $S(p)$ is small. The absolute magnitude of $S(p)$ is generally not interpreted; only the relative change as a function of p is. In this way, the structure factor serves as a signal of correlations in a configuration at a given periodicity.

It is widely held that the excess entropy E serves as a general-purpose measure of a system's structure, regularity, or memory; for recent reviews, see Refs. [2,28,29]. The excess entropy provides a quantitative measure of structure that may be applied to any 1D symbolic string. In Refs. [27,32,33], we argued that E may be viewed as an effective order parameter for 1D spin systems. In particular, we showed that the excess entropy is sensitive to periodic structure at any period, whereas structure factors, by construction, are sensitive to ordering at only a single spatial period. We shall return to this point below and show that the same general claims hold in two dimensions as well.

III. TWO-DIMENSIONAL ENTROPY, ENTROPY DENSITY, AND EXCESS ENTROPY

A. Generalizing to higher spatial dimensions

Below we discuss how to extend the 1D analysis outlined above to apply to spatial patterns in two and higher dimensions. Before launching into definitions and formalism, we sketch some of the philosophy and intuitions that motivate the path we take and highlight some of the general issues that arise as one moves from 1D to 2D systems.

Patterns in two dimensions are fundamentally different from those in one dimension. For example, in one dimension a natural way to scan a configuration exists: left-to-right, say. That is, each local variable is indexed in a well-defined order. (The information-theoretic measures discussed in the preceding section have the same values regardless of whether the 1D configuration is scanned left-to-right or right-to-left.)

The 1D approach simply does not generalize to 2D in a unique natural way. One might be tempted to scan or parse a 2D configuration by taking a particular 1D path through it. One would then apply 1D measures of randomness and structure to the sequences thus obtained. For example, in Refs. [34,35], a space-filling curve is used to parse a 2D configuration and, from this, the entropy density of the configuration is estimated.

While the 1D-path method does yield the correct entropy density, it is also clear that it projects additional spurious structure onto the configuration. By snaking through the lattice, it is inevitable that sites, adjacent in the 2D lattice, occur far apart in the 1D sequence. As a result, long-range correlations appear in the latter. Thus, a 1D excess entropy (or any other 1D measure of structural complexity) adapted in this way will capture not only properties of the 2D configuration, but also properties of the path. Except in special cases and with appropriate prior knowledge, it does not appear possible to disentangle these two distinct sources of apparent structure. These and related difficulties with the 1D approach have been discussed in some detail in, for example, Refs. [22,27,36].

Here, we seek an alternative to understanding a 2D pattern by parsing it into 1D strings. We are immediately faced with a problem, however. There is a unique, complete ordering of the connected, nested subsets of a 1D lattice such that the conditional entropies of the target spin, conditioned on this sequence of subsets, are monotone decreasing. It is this ordering that makes Eq. (10) unambiguous and unique in 1D. In contrast, connected, nested subsets of a 2D lattice that have this monotonic property are not unique. This is a direct consequence of the topological differences between one- and two-dimensional lattices. We shall see that this lack of uniqueness introduces ambiguity in extending Eq. (10) to two dimensions; specifically, there is no natural unique expression for the excess entropy in two dimensions.

B. Entropy density

The entropy density in two dimensions is defined in the natural way. Consider an infinite 2D square lattice of random

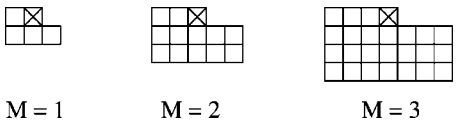

FIG. 1. Neighborhood templates for 2D conditional entropies. The target spin is denoted with an X.

variables S_{ij} whose values range over the finite set $\mathcal{A}$. Assuming that the variables are translationally invariant, the 2D entropy density is given by

$$h_\mu = \lim_{N,M\to\infty} \frac{H(N,M)}{NM}, \tag{18}$$

where $H(N,M)$ is the Shannon entropy of an $N\times M$ block of spin variables. This limit exists for a translationally invariant system, provided that the limits are taken in such a manner that the ratio N/M remains constant and finite.

Is there a way to reexpress the 2D entropy density of Eq. (18) as the entropy of a target variable conditioned on a block of neighboring variables, analogous to Eq. (9)? This question was, to the best of our knowledge, first answered in the affirmative by Alexandrowicz in the early 1970s [37,38]. Meirovitch [39,40] and later Schlijper and co-authors [41,42] extended and applied Alexandrowicz's work. These methods have also been discovered independently by Eriksson and Lindgren [43,44] and Olbrich *et al.* [45]. Here, we briefly summarize the central result and adapt it to our needs.

The most general approach to the conditional entropy in two dimensions proceeds as follows. Let $h_\mu(M)$ denote the Shannon entropy of the target spin conditioned on a 2D neighborhood template of $2M(M+1)$ spins. Arrange the spin template in an $(M+1)\times(2M+1)$ rectangle with the target spin in the center of the rectangle's top row and with the top, rightmost M spins deleted from the template. A sequence of neighborhood templates of this type is shown in Fig. 1. For example, $h_\mu(3)$ is the entropy of the target spin (denoted by an X) conditioned on all the other spins in the rightmost template of Fig. 1. The 2D entropy density h_μ may then be shown to be equal to [43,44,46,47]

$$h_\mu = \lim_{M\to\infty} h_\mu(M). \tag{19}$$

If it is known that the interactions between spins are of finite range, then one only needs to use a shape as thick as the interaction range [37,38,41–43]. For example, in the following section, we consider a 2D Ising model with nearest- and next-nearest-neighbor (NNN) interactions. In this case, one uses a strip with a thickness of two lattice sites; see Fig. 2.

We now slightly modify the definition of the template-size parameter M in the conditional single-site entropy $h_\mu(M)$ so as to apply to the scenario in Fig. 2. The cell numbers in this figure indicate the order in which individual sites are added to the neighborhood template. For example, $h_\mu(3)$ now will denote the entropy of the target spin S_{00} conditioned on the three spins labeled 1 (S_{-10}), 2 (S_{01}), and 3 (S_{-11}); that is,

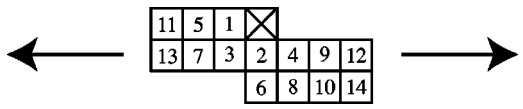

FIG. 2. Target spin (X) and neighborhood templates for conditional entropies used in our study of the 2D NNN Ising model. The cell numbers indicate the order in which the sites are added to the template. For more discussion, see text.

$$h_\mu(3) = H[S_{00}|S_{-10},S_{01},S_{-11}]. \tag{20}$$

In the $M\to\infty$ limit, the new $h_\mu(M)$ still goes to the entropy density, as in Eq. (19). It is also not hard to see that this convergence must be monotonic

$$h_\mu(M) \leqslant h_\mu(M'), \quad M > M'. \tag{21}$$

This is a direct consequence of the fact that conditioning cannot increase entropy [30].

Two remarks about the neighborhood template in Fig. 2 are in order. First, the strip needs to be two sites thick since the system explored below has interactions that extend across two lattice sites. In this case, a strip with a thickness of two sites shields one-half of the lattice from the other. In the limit that the strip is infinitely long in the horizontal direction, then the probability distribution of the target spin is independent of the values of the spins beneath the strip [47].

Second, at first blush, the numbering scheme in Fig. 2 appears ambiguous. Spins are added to the template in order of increasing Euclidean distance from the target spin. For example, spin 10 is a Euclidean distance $2\sqrt{2}$ from the center spin, whereas spin 11 is a distance of 3. Since $2\sqrt{2}<3$, one adds on spin 10 before 11. When there is a tie, one adds the leftmost spin. For example, spins 3 and 4 are the same Euclidean distance from the center spin; spin 3 comes before 4 since it is to the left.

Of course, one can use alternative ordering schemes, such as adding spins in a widening spiral or some other geometric pattern. These choices do not change the result in Eq. (19), since this is a statement about what happens in the limit that an arbitrarily large number of spins have been added to the template. However, looking ahead, the order in which spins are added can affect the convergence form of the 2D excess entropy—the 2D analog of E_C of Eq. (10).

As noted above, the ambiguity in how the neighborhood template of conditioning variables grows is a direct result of the fact that a 2D lattice does not specify a strict ordering of its elements in the way that a 1D sequence does. Rather, a 2D lattice specifies a partial ordering of its elements. Thus, there will always be ''ties'' in the sense just mentioned, and so there is no unique natural way to add on the spins one-by-one based on an ordering of subsets of spin blocks. See Ref. [48] for a detailed discussion of this, albeit in a slightly different context.

Finally, the conditional Shannon entropy method for calculating the entropy density h_μ is well known and has been successfully applied to a number of different systems; including the 2D Ising model on square [42] and simple cubic

lattices [40], the $q=5$ 2D Potts model [42], a 2D hard-square lattice gas [49], the three-dimensional fcc Ising antiferromagnet [50], coupled map lattices [45], Gaussian random fields [51], polymer chain models [52], and network-forming materials [53]. Quite recently, Meirovitch [52] estimated the entropy for the 2D Ising ferromagnet. Remarkably, his results have only a 0.01% relative error at the critical temperature, where one might expect the conditional entropy form to overestimate the entropy density due to long-range correlations missed by finite-size templates.

C. Excess entropy in two dimensions

We now turn to the question of how to extend excess entropy to more than one dimension. We consider three possible approaches to excess entropy in two dimensions. For each, we begin with one of the three different forms for the 1D excess entropy.

First, consider the convergence excess entropy E_{C}, as defined in Eq. (10). In the preceding section, we defined a sequence of 2D entropy density estimates $h_\mu(M)$ that converges from above to the entropy density h_μ. We can sum these entropy density overestimates to obtain the 2D convergence excess entropy:

$$E_{\mathrm{C}} \equiv \sum_{M=1}^{\infty} [h_\mu(M) - h_\mu]. \tag{22}$$

We shall see that this form of the excess entropy is, like its 1D cousin, capable of capturing the structures or correlations present in a 2D system. Note that this definition can depend on the order in which spins are added on to the template and, as discussed in the preceding section, there is no unique ordering to use to determine the sequence in which to add sites. Nevertheless, our investigations have shown that any reasonable choice for ordering yields an E_{C} that behaves qualitatively the same as that defined in Eq. (22).

The mutual information form E_{I} of the excess entropy, defined in Eq. (13), can naturally be extended by considering the mutual information between two adjacent infinite half-planes:

$$E_{\mathrm{I}} \equiv \lim_{M,N\to\infty} I\left[\begin{array}{c} \uparrow \\ N \\ \downarrow \end{array}\right.\!\!\begin{array}{c} \longleftarrow M \longrightarrow \\ \square\square\square\square \\ \square\square\square\square \end{array}\;;\;\begin{array}{c} \longleftarrow M \longrightarrow \\ \square\square\square\square \\ \square\square\square\square \end{array}\!\!\left.\begin{array}{c} \uparrow \\ N \\ \downarrow \end{array}\right]. \tag{23}$$

As in Eq. (13), it is understood that the two semi-infinite planes are adjacent.

Finally, one may also develop an expression for 2D subextensive excess entropies by considering how $H(M,N)$ grows with M and N. In analogy to Eq. (11), we define three subextensive excess entropies via

$$H(M,N) = H\left[\begin{array}{c} \longleftarrow M \longrightarrow \\ \square\square\square\square \\ \square\square\square\square \end{array}\!\!\begin{array}{c} \uparrow \\ N \\ \downarrow \end{array}\right] \tag{24}$$

$$\sim E_{\mathrm{S}} + E_{\mathrm{S}}^x M + E_{\mathrm{S}}^y N + h_\mu MN. \tag{25}$$

Note that in an isotropic system, such as that considered below, $E_{\mathrm{S}}^x = E_{\mathrm{S}}^y$. We shall not consider these forms for the 2D excess entropy here, opting instead to focus on E_{C} and E_{I}.

IV. RESULTS

A. Next-nearest-neighbor Ising systems

To test the behavior of the different forms of the excess entropy, we estimated E_{I} and E_{C} numerically for a standard system: the 2D spin-$\frac{1}{2}$ Ising model with nearest-neighbor (NN) and next-nearest-neighbor (NNN) interactions. We chose this system since it is rich enough to exhibit several distinct structures and due to its broad familiarity. Its Hamiltonian, $\mathcal{H}$, is given by

$$\mathcal{H} = -J_1 \sum_{\langle ij,kl\rangle_{\mathrm{nn}}} S_{ij}S_{kl} - J_2 \sum_{\langle ij,kl\rangle_{\mathrm{nnn}}} S_{ij}S_{kl} - B\sum_{ij} S_{ij}, \tag{26}$$

where the first (second) sum is understood to run over all NN (NNN) sets of spins. Each spin S_{ij} is a binary variable: $S_{ij} \in \{-1,+1\}$. The lattice consists of $N\times N$ spins; the spatial indices on spin variables run from 0 to $N-1$. We shall view the spins as dimensionless variables and the coupling constants J_1 and J_2 as dimensionless parameters.

We estimated the structure factors $S(1)$, $S(2)$, and $S(4)$ with Eq. (17) by directly measuring the frequency of occurrence of $s_i s_j$ and s in spin configurations generated by a Monte Carlo simulation that used a standard single-site Metropolis algorithm on a lattice with periodic boundary conditions. That is, we sampled configurations with the canonical distribution $e^{-\mathcal{H}(c)/T}$, where $\mathcal{H}(c)$ is the energy of the configuration c and T is the temperature. We used a lattice of 48×48 spins. Since we are not interested here in extracting the system's critical properties, there is no need to go to larger system sizes.

We estimated E_{C} and E_{I} from block probabilities by observing the frequency of spin-block occurrences. To estimate E_{C}, we used a template containing 15 total spins, as shown in Fig. 2, and marginals of this distribution for smaller template sizes. To estimate E_{I} we calculated the mutual information of two adjacent 2×4 spin blocks. For each J_1 value, we ran our Monte Carlo simulation for up to 2×10^5 Monte Carlo time steps (2×10^6 for $J_1 < -1.5$) and then took data every 20 time steps for 2×10^4 time steps. One Monte Carlo time step corresponds to trying to flip, on average, each spin in the lattice one time. We thus sampled $\approx 2\times 10^6$ template configurations. For comparison, note that there are at most (in the highly disordered regime) $2^{16} \approx 3\times 10^4$ possible configurations in a template of 16 spins.

B. Excess entropy detects periodic structure

Our results are shown in Fig. 3. The temperature was held at $T=1.0$, the external field at $B=0.0$, and the next-nearest-neighbor coupling at $J_2=-1.0$. Figure 3 shows $S(1)$, $S(2)$, $S(4)$, E_{I}, and E_{C}, as a function of $J_1 \in [-4.0,4.0]$. For all J_1 values, the temperature is relatively small compared to the average energy per spin. And so, the configurations sampled are typically the ground state with a few low-energy excitations.

As J_1 is increased, the system moves through parameter regimes in which there are significant correlations of periods

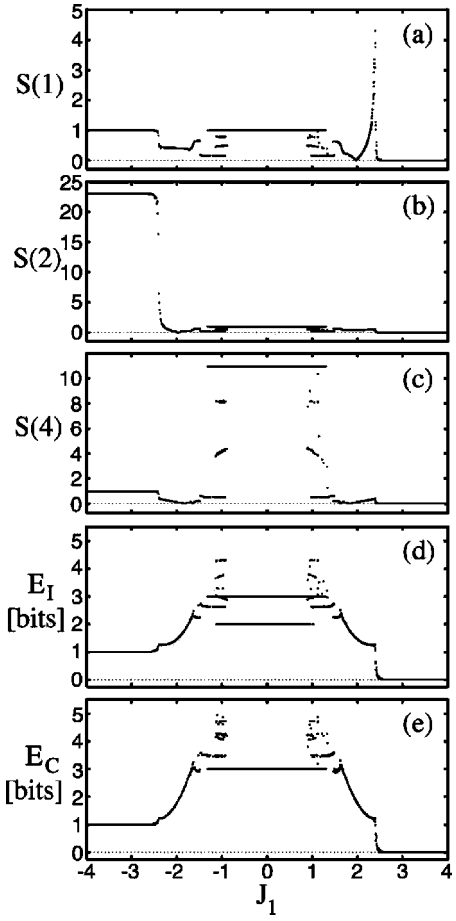

FIG. 3. Structural changes in the the 2D NNN Ising model as a function of NN coupling J_1 as revealed by structure factors (a) $S(1)$, (b) $S(2)$, and (c) $S(4)$, and excess entropies (d) E_I (mutual information) and (e) E_C (convergence). The temperature was fixed at $T=1.0$ and J_2 was held at -1.0 as the NN coupling was swept from $J_1=-4.0$ to $J_1=4.0$ in steps of $\delta J_1=0.01$, except near the $S(1)$ spike at $J_1\approx 2.5$ where $\delta J_1=0.005$. We performed at least five different runs at each J_1 in the range $|J_1|\leqslant 1.15$. Note the different scales on the vertical axes: the excess entropies are measured in bits of apparent memory; the structure factors and J_1 are dimensionless. For more discussion, see text.

2, 4, and 1. This is seen, for example, in the behavior of the various structure factors; the structure factors selected correspond to periods of 2, 4, and 1 lattice sites.

Physically, when J_1 is large in magnitude and negative, the tendency for nearest neighbors to anti-align dominates and the system's ground state is antiferromagnetic: a checkerboard pattern consisting of alternating up and down spins. This pattern has a spatial period of 2. Not surprisingly, the period-2 structure factor $S(2)$ in Fig. 3(b) shows a strong signal in this low-J_1 regime.

When J_1 is near zero, the NN interactions are negligible compared to the NNN interactions. Thus, each spin orients opposite its four next-nearest neighbors, while disregarding its four nearest neighbors. The result is that the lattice effectively decouples into four noninteracting sublattices. On each of these sublattices, the spins alternate in sign, resulting in a ground state with spatial period 4. Note that the period-4 structure factor $S(4)$ in Fig. 3(c) has a large value near $J_1=0$, indicating this period-4 ordering.

As J_1 is increased from 0, the tendency for the spins to align grows stronger. Eventually, this NN interaction overwhelms the NNN interactions and the entire lattice starts to align. This is the familiar paramagnet-ferromagnet transition. Above $J_1\approx 2.5$, the system acquires a net magnetization; there is now an unequal number of up and down spins, whereas below $J_1\approx 2.5$ there are always, on average, equal numbers of up and down spins. This transition is signaled by the distinct spike in the period-1 structure factor $S(1)$ near $J_1\approx 2.5$ and $S(1)$'s vanishing at larger J_1. [The magnetic susceptibility χ diverges at the critical point of a ferromagnet-paramagnet transition. Since $\chi\propto S(1)$, one expects to see a spike in $S(1)$ near this transition, where the system acquires a nonzero magnetization.]

In Figs. 3(d) and 3(e) we plot the mutual-information excess entropy E_I and the convergence excess entropy E_C versus J_1 over the same parameter range. In the large and negative J_1 regime, $E_I=E_C=1$ bit, indicating that there is one bit of information stored in the configurations. The configurations have a simple structure (alternating up and down spins) and the magnitude of E gives the information needed to specify the spatial phase of the period-2 configurations. When J_1 is large and the system undergoes the transition to ferromagnetic ordering, $E_I=E_C=0$, since the configurations consist of all aligned spins, and there is no spatial information or structure in them. In the intermediate regime ($J_1\approx 0$), E_I and E_C are markedly larger, indicating that the system is more structured than elsewhere. We will return shortly to discuss in detail what the values of E_I and E_C mean.

Note that each excess entropy is sensitive to correlations at *all* periodicities, despite the fact that each is merely a single unparametrized function. In contrast, the structure factors $S(p)$ are a one-parameter family of functions that must be tuned *a posteriori* to find relevant periodic structure. That is, the period-1 structure factor $S(1)$ detects only the period-1 correlations near $J_1=2.5$. Moreover, $S(1)$ is unable to distinguish between the period-2 and period-4 orderings at $J_1<-3.0$ and $J_1\approx 0$, respectively; $S(1)\approx 1$ for both period-2 and period-4 configurations.

Since the excess entropy E is a single unparametrized function sensitive to structure of any periodicity, it is a more general measure of structure and correlation than the structure factors $S(p)$. Conversely, $S(p)$ is somewhat myopic. By considering only two-point correlations modulated at a selected periodicity p, $S(p)$ misses structure that is either aperiodic or that is due to more than two-spin correlations. In fact, E is even more sensitive and general than these observations indicate.

C. *E* distinguishes structurally distinct ground states

Looking closely at the mutual-information excess entropy E_I near $J_1=0$ in Fig. 3(d), one notices that the curve splits into two in the $|J_1|<1.0$ region. This can be seen more clearly in Fig. 4, in which we plot E_I versus J_1 in this region. We sampled the NN coupling J_1 every 0.01 and performed at least five different runs at each J_1 value. Sometimes E_I

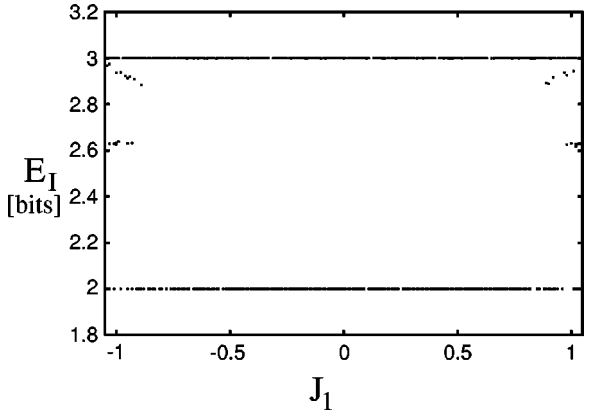

FIG. 4. The mutual-information excess entropy E_I showing the existence of multiple period-4 ground states.

$=3.0$ bits, whereas for other trials $E_I=2.0$ bits. Why are there two different values for E_I on different runs? And why, in contrast, is the period-4 structure factor $S(4)$ the same for all runs?

The answer is simple: there are multiple structurally distinct ground states. The three possible ground-state configurations are shown in Fig. 5. Note that for each ground state, all NNN pairs of sites have opposite spin values, thus minimizing the system's energy. Note also that each ground state is identical if one considers only a horizontal or vertical slice; the repeating pattern of two up spins followed by two down spins is the same.

After a long transient time, the system usually settles into one of these three states. A boundary defect between two different ground states has an energy cost associated with it. As such, most boundaries are eventually destroyed. Incidentally, the dynamics through which this removal of boundary defects occurs is rather subtle and can be very long lived. For example, a boundary between left and right diagonal phases costs more than a boundary between the checkerboard and one of the striped patterns. As a result, when the two different striped phases come close, the checkerboard pattern emerges between them, pushing the stripe boundaries away from each other. Moreover, as the temperature approaches zero, we observe that there are times when the ground state is simply not found via single-flip Metropolis Monte Carlo dynamics. Similar phenomena have been observed in other antiferromagnetic Ising models; for recent work, see Refs. [54–56].

In any event, a straightforward calculation shows that $E_I=3$ bits for the checkerboard configuration of Fig. 5(a),

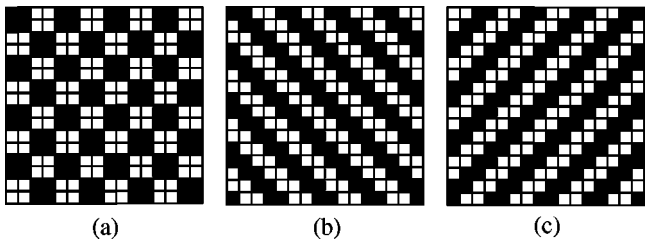

FIG. 5. The three ground states for $J_1\approx 0$ and $J_2<0$: (a) checkerboard, (b) left-diagonal stripe, and (c) right-diagonal stripe.

whereas $E_I=2$ bits for the two striped phases. Similar calculations show that $E_C=3$ bits for both the checkerboard and striped ground states. Note, however, that $S(4)$ is the same for all three ground states. By construction, $S(4)$ measures only two-spin statistics obtained by considering correlations along a horizontal or a vertical direction. And so, the three ground states are the same if one considers only isolated horizontal or vertical slices; every slice consists of a repeating pattern of two up spins followed by two down spins. Of course, one can adapt the definition of $S(p)$ to account for the diagonal striped phases, but this simply begs the question of discovering the intrinsic patterns in the first place.

Near $|J_1|=1$, notice that E_I and E_C occur in plateaus between 2 and 3 bits and above. This indicates that the system has settled into one of several more structured metastable states consisting of mixtures of the three ground states.

In summary, we see that the mutual information excess entropy E_I is capable of distinguishing between patterns that are not distinct according to the structure factors $S(p)$. In fact, we initially did not anticipate the two striped ground states, glibly assuming that the only ground state is the checkerboard. Our results for E_I, which we initially found confusing, led us to examine the configurations more closely and to detect the distinct ground state structures. This, in turn, led us to notice the rich dynamics of the configurations as they wend their way towards one of the three ground states. In short, these structural subtleties would have been missed entirely had we relied solely on the structure factors.

V. DISCUSSION AND CONCLUSION

We introduced three extensions of the excess entropy that apply to two-dimensional configurations. Each excess entropy expression is based on a different way of viewing the one-dimensional excess entropy: the convergence excess entropy E_C measures the manner in which finite-template entropy density estimates converge to their asymptotic value; the subextensive excess entropy E_S is related to the subextensive forms of the block entropy $H(M,N)$; and the mutual information excess entropy E_I is defined as the mutual information between two halves of a configuration.

Applying two of these measures, E_C and E_I, to the NNN Ising model, we have seen that these quantities capture the structural changes this system undergoes as its parameters are varied. In contrast, the structure factors are sensitive to periodic ordering of a particular period. Moreover, our results show that the information excess entropy E_I clearly distinguishes between two period-4 ground states, whereas the period-4 structure factor is simply incapable of making such a structural distinction. Finally, the values that the excess entropies take on are interpretable and give a quantitative measure of the amount of structure in the system.

The picture that emerges, then, is that the various two-dimensional excess entropies are general-purpose measures of two-dimensional structure. This is not to suggest that the excess entropy replace structure factors or, more generally, Fourier analysis. We view the excess entropy not in competition with Fourier analysis, but complementary to it; the excess entropy is designed to answer a different set of ques-

tions than those addressed by Fourier components. For example, it has long been appreciated in dynamical systems that power spectral analysis is of little help in revealing the geometry of a chaotic attractor [57]. Analogously, spectral decomposition, typically, will say little about how difficult it is to learn or synchronize to a pattern.

Clearly, however, there is much more work to be done to develop a thorough, well understood methodology for two-dimensional patterns. One possible approach builds on Refs. [4,28,58] which take a systematic look at entropy growth and convergence by using a discrete calculus. This work places several complexity measures within a common framework and leads to new measures of structure. From the study presented above, we conclude that a similar analysis in two dimensions, using a two-dimensional discrete calculus, holds great promise. Another area for future research concerns developing relationships between measures of complexity of a pattern and the difficulty of learning or synchronizing to it [2–4,28].

There are also, of course, a host of additional statistical mechanical systems, each with its own range of distinct structures, that should be similarly analyzed. Calculating excess entropies for them will facilitate developing our understanding of the behavior of these different quantities and may even lead to discovering novel structural properties. A natural choice is calculating the behavior of E near the critical temperature, extracting critical exponents, and relating these exponents to others for the well studied nearest-neighbor Ising model. It will also be of interest to calculate the various excess entropy forms for noisy Sierpinsky carpets and the like; this will allow for direct comparison with calculations of the measures of inhomogeneity put forth in Refs. [11,12].

Ultimately, these different measures of structure—those presented here and those developed by other authors—will be judged not solely by their ability to shed light on existing well understood model systems such as the NNN Ising model considered here. Instead, the broader concern is how to use these information-theoretic quantities to capture structure and patterns in systems that are less well understood. Equally important is the question of establishing relationships between information-theoretic measures of structural complexity and other quantities, including: physical measures of structure and correlation; computation-theoretic properties; and the difficulty of learning a pattern.

ACKNOWLEDGMENTS

We thank Kristian Lindgren, Susan McKay, and Karl Young for helpful discussions. This work was supported by the Santa Fe Institute under the Computation, Dynamics, and Inference Program via SFI's core grants from the National Science and MacArthur Foundations. Direct support was provided from DARPA under Contract No. F30602-00-2-0583. D.P.F. thanks the Department of Physics and Astronomy at the University of Maine for their hospitality. The Linux cluster used for the simulations reported here was provided by Intel Corporation through its support of SFI's Network Dynamics Program.

[1] J.P. Crutchfield, Physica D **75**, 11 (1994).
[2] W. Bialek, I. Nemenman, and N. Tishby, Neural Comput. **13**, 2409 (2001).
[3] I. Nemenman, Ph.D. thesis, Princeton University, 2000 (unpublished).
[4] J.P. Crutchfield and D.P. Feldman, Adv. Complex Syst. **4**, 251 (2001).
[5] L. Dębowski, in *Nonextensive Entropy: Interdisciplinary Applications*, edited by C. Tsallis and M. Gell-Mann (Oxford University Press, Oxford, 2003).
[6] D.P. Varn, G.S. Canright, and J.P. Crutchfield, Phys. Rev. B **66**, 174110 (2002).
[7] C. Andraud, A. Beghdadi, and J. Lafait, Physica A **207**, 208 (1994).
[8] C. Andraud, A. Beghdadi, E. Haslund, R. Hilfer, J. Lafait, and B. Virgin, Physica A **235**, 307 (1997).
[9] C.D.Van Siclen, Phys. Rev. E **56**, 5211 (1997).
[10] R. Piasecki, Physica A **277**, 157 (2000).
[11] R. Piasecki, Surf. Sci. **454-456**, 1058 (2000).
[12] R. Piasecki, M.T. Martin, and A. Plastino, Physica A **307**, 157 (2002).
[13] A.N. Pavlov, W. Ebeling, L. Molgedey, A.R. Ziganshin, and V.S. Anishchenko, Physica A **300**, 310 (2001).
[14] J.L. McCauley, Phys. Rep. **189**, 225 (1990).
[15] C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems* (Cambridge University Press, Cambridge, 1993).
[16] K. Young and J.P. Crutchfield, Chaos, Solitons Fractals **4**, 5 (1993).
[17] K. Lindgren, C. Moore, and M.G. Nordahl, J. Stat. Phys. **91**, 909 (1998).
[18] Y.A. Andrienko, N.V. Brilliantov, and J. Kurths, Eur. Phys. J. B **15**, 539 (2000).
[19] E.M. Sal'nikova and L.M. Martyushev, Tech. Phys. Lett. **27**, 301 (2001).
[20] J.P. Crutchfield and N.H. Packard, Physica D **7**, 201 (1983).
[21] R. Shaw, *The Dripping Faucet as a Model Chaotic System* (Aerial Press, Santa Cruz, 1984).
[22] P. Grassberger, Int. J. Theor. Phys. **25**, 907 (1986).
[23] K. Lindgren and M.G. Norhdal, Complex Syst. **2**, 409 (1988).
[24] W. Li, Complex Syst. **5**, 381 (1991).
[25] D. Arnold, Complex Syst. **10**, 143 (1996).
[26] D.P. Feldman and J.P. Crutchfield, Phys. Lett. A **238**, 244 (1998).
[27] D.P. Feldman, Ph.D. thesis, University of California, Davis, 1998 (unpublished).
[28] J.P. Crutchfield and D.P. Feldman, Chaos **13**, 25 (2003).
[29] W. Ebeling, Physica D **109**, 42 (1997).
[30] T.M. Cover and J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[31] T. Schürmann and P. Grassberger, Chaos **6**, 414 (1996).
[32] J.P. Crutchfield and D.P. Feldman, Phys. Rev. E **55**, R1239 (1997).

[33] D.P. Feldman and J.P. Crutchfield, Santa Fe Institute Report No. 98-04-026, 1998 (unpublished).
[34] A. Lempel and J. Ziv, IEEE Trans. Inf. Theory **32**, 2 (1986).
[35] D. Sheinwald, A. Lempel, and J. Ziv, IEEE Trans. Inf. Theory **38**, 341 (1990).
[36] E. Soljanin, IEEE Trans. Inf. Theory **48**, 1344 (2002).
[37] Z. Alexandrowicz, J. Chem. Phys. **55**, 2765 (1971).
[38] Z. Alexandrowicz, J. Stat. Phys. **14**, 1 (1976).
[39] H. Meirovitch, Chem. Phys. Lett. **45**, 389 (1977).
[40] H. Meirovitch, J. Phys. A **16**, 839 (1983).
[41] A.G. Schlijper, A.R.D. van Bergen, and B. Smit, Phys. Rev. A **41**, 1175 (1990).
[42] A.G. Schlijper and B. Smit, J. Stat. Phys. **56**, 247 (1989).
[43] K.-E. Eriksson and K. Lindgren, Chalmers University of Technology and University of Göteborg Technical Report, 1989 (unpublished).
[44] K. Lindgren, in *Beyond Belief: Randomness, Prediction and Explanation in Science*, edited by J. L. Casti and A. Karlqvist (CRC Press, Boca Raton, 1991), pp. 88–109.
[45] E. Olbrich, R. Hegger, and H. Kantz, Phys. Rev. Lett. **84**, 2132 (2000).
[46] A.G. Schlijper, Phys. Rev. B **27**, 6841 (1983).
[47] S. Goldstein, R. Kuik, and A.G. Schlijper, Commun. Math. Phys. **128**, 469 (1990).
[48] B. Baumgartner, J. Phys. A **35**, 3163 (2002).
[49] H. Meirovitch, J. Stat. Phys. **30**, 681 (1983).
[50] H. Meirovitch, Phys. Rev. B **30**, 2866 (1984).
[51] S. Marcelja, Physica A **231**, 168 (1996).
[52] H. Meirovitch, J. Chem. Phys. **1111**, 7215 (1999).
[53] R.L.C. Vink and G.T. Barkema, Phys. Rev. Lett. **89**, 076405 (2002).
[54] V. Spirin, P. Krapivsky, and S. Redner, Phys. Rev. E **63**, 036118 (2001).
[55] V. Spirin, P. Krapivsky, and S. Redner, Phys. Rev. E **63**, 016119 (2001).
[56] F. Vazquez, P.L. Krapivsky, and S. Redner, e-print arXiv-01 cond-mat/0209445.
[57] J.D. Farmer, J.P. Crutchfield, H. Froehling, N.H. Packard, and R.S. Shaw, Ann. N.Y. Acad. Sci. **357**, 453 (1980).
[58] P.-M. Binder and J.A. Plazas, Phys. Rev. E **63**, 065203(R) (2001).